Linked Data as enabler for Open Data Ecosystems

Miel Vander Sande, Anastasia Dimou, Pieter Colpaert, Erik Mannens, Rik Van de Walle
Ghent University - iMinds - Multimedia Lab
{miel.vandersande, anastasia.dimou, pieter.colpaert, erik.mannens, rik.vandewalle}@ugent.be

Introduction

The Open Data movement is widely spreading. Its open governmental data perspective is driven by the fundamental idea that governments have the right and duty to publish the public data. To this end, there is response from a lot of organizations, not only governments, to release the data they collect. However, what happens after the data are published?

Till now, it is the data publisher's responsibility to maintain the Open Data and keep them up-to-date. But how easy is it for the data publisher to maintain the data and how tolerant is the current publishing method to allow the publisher to apply changes? What if data consumers start to contribute? How are intellectual property rights going to be managed when third parties contribute to the data?

For the time being, data consumers have not been able to provide feedback or contribute to the released data. Open Data are still published as a one-way channel between the data publisher and the data consumer. The data consumer can access the available Open Data, but neither can he send feedback directly (e.g., comments, flag errors, etc.) nor can he edit the data in an attempt to update, extend, or correct the existing ones. If so, one needs to use an alternative channel (e.g., e-mail). This leads to a complicated feedback procedure, potential loss in data, and rather inefficient data management.

In order to overcome these, we introduce the idea of Linked Data as an enabler for Open Data Ecosystem. Linked Data can support the implementation of the feedback loop and lead to full Data Cycles in their ecosystem.

Ecosystems and their requirements

Since the feedback loop is currently missing, the Data Cycles in this ecosystem are disturbed [1]. A complete Data Cycle demands that, after the data are released, data consumers are not only able to view the data, but they can also contribute and provide their feedback. They are able to edit the data in order to apply corrections or update them so as to contribute to their maintenance. Also, they can provide their comments or annotate the data. Therefore, providing this feedback possibility is of crucial importance, since it enables the data consumers to remain actively involved and maximizes the uppermost goal of public participation. Only then the potential benefits as posed by Neelie Kroes can be achieved [2].

We believe that the feedback loop is required to complete the missing Data Cycles when publishing Open Data. To this end, two challenges need to be achieved: a uniform Read-Write interface in the front-end and a powerful Data Management system in the back-end. It is required that the Web Interface is able to provide means to view the published Open Data and edit them, while the underlying Data Management System should be more focused on handling the data storage and publication.

Linked Data as enabler

In this section, we propose the Linked Data as a promising technology able to fulfill the Open Data Ecosystem's requirements and become the basis for a Read–Write platform. Linked Data technologies are considered to be in the position to tackle with the demands raised upon shifting from a one-way channel to a homogeneous bidirectional one. They can successfully implement the co-ownership -different contributors' right on the published data, versioning and provenance concepts raised. At the same time, they are able to achieve the challenge of supporting advanced interoperability between the Web Clients and the Data Management System. In this context, the W3C has already published a working draft of a Linked Data Platform, which describes a set of practices for a Read-Write Linked Data architecture [3].

Read – Write Web Interface

To ensure that the data consumers are able to interact with Open Data, easy to access and easy to use Web Clients are required. Relying on REST for the communication between the client and the underlying data core, interoperability is achieved. This means that on top of the Data core, different web applications can be built. Furthermore, building upon HTTP and web standards, a uniform interface can be implemented which can fully interoperate with the Linked Data technologies. From now on, data consumers are able to interact with more than one dataset at once: they can send queries against different datasets and collect aggregated data in different visualization formats.

Data consumers also interact with the Open Data by editing the data. Their feedback can have the form of a new version submission of the data or part of the data, providing corrections or updates. On the other side, the data consumers' contribution may target to the data's metadata, namely providing comments or annotations.

Data Management

In the back-end, the data storage and publication is implemented. The data are identified by meaningful HTTP URIs which correspond to each resource's RDF representation. Therefore, it is machine readable, understandable and browsable, augmenting its potential as it enables a serious increase in automation and handling of bigger and more heterogeneous data volumes.

The Linked Data system can support the retrieval and querying of the data requested by the Web Clients in order for the data consumers to view them. A Web Client can query not only the data itself but also its metadata (e.g., the license, the provenance, the mean time until update). But as the data consumers become data participants and as they are able to contribute to the data directly at the source, the corresponding Linked Open Data versioning, trust, provenance and ownership aspects come into discussion as they need to be handled by the underlying Data Management system. The Linked Data paradigm can support these emerging issues too.

The Data Management system can identify the data consumer as the Web Client provides this information (e.g., the user's WebID) bound to the data consumer's contribution. The data's version is enriched with metadata related to the user's identity, thus the provenance of the data can be tracked. As provenance is the key to reach transparency, it is of crucial importance to guarantee the data's legitimacy by keeping track of the data's provenance.

On a Linked Data based system, a distributed triple version control system is used to keep track of the different versions of the data and its metadata. Based on this system, the different contributions are summarized. As each participant's contribution is known, their ownership right can be calculated, ensuring co-ownership. Furthermore, their edition contributes to the overall improvement of the data quality. A data participant's contribution can be an update to the data itself, but it can also be a contribution on the data's metadata, for instance a comment or an annotation.

Finally, the Linked Data technology can handle the data consumers' request for aggregated data by returning to the data consumer only the requested part of the data. At the same time, the system profits of the usage statistics, identifying further option for segmentation and interlinkage with other data enhancing its data searchability and discoverability. Finally, the Linked Data based Management System does not allow the aggregated data to leak out of the Open Data Cycle, but can support the incorporation of the derived processed information back in the Open Data Cycle, taking always into consideration the license metadata of each dataset.

Conclusions

At this document, we point the broken Data Cycles on current Open Data Ecosystem due to the missing feedback loop. As long as there is no out-of-the-box solution, we suggest that Linked Data technologies can be the starting point for a Read – Write platform which is able to address the current lack of a homogeneous feedback loop and handle, at the same time, the emerging co-ownership, versioning and provenance issues. The proposed solution comprises a uniform Web Interface based on HTTP and a powerful Data Management system based on Linked Data. The W3C Working Draft of a Linked Data Platform gives a good set of practices. Considering the aforementioned, we intend to bring into discussion whether a corresponding set of practices for a Linked *Open* Data Platform tackling the above mentioned issues, would be needed.

References

1. Building the Open Data Ecosystem, 2011, http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/ (last accessed on 03/01/2013)
2. Neelie Kroes. Digital agenda: Commission's open data strategy, questions & answers, 2011, http://europa.eu/rapid/press-release_MEMO-11-891_en.htm?locale=en (last accessed on 03/01/2013).
3. Linked Data Platform 1.0 Working Draft, 2012, http://www.w3.org/TR/ldp/ (last accessed on 03/01/2013)